DOCUMENT RESUME

ED 477 060 IR 021 788

AUTHOR Maini, Gaurav; Leggett, John J.; Ong, Teongjoo; Wilson, Hugh D.; Reed, Monique D.; Hatch, Stephan L.; Dawson, John E.

TITLE Enhancing Scientific Practice and Education through Collaborative Digital Libraries.

PUB DATE 2002-06-00

NOTE 6p.; In: ED-MEDIA 2002 World Conference on Educational Multimedia, Hypermedia & Telecommunications. Proceedings (14th, Denver, Colorado, June 24-29, 2002); see IR 021 687.

AVAILABLE FROM Association for the Advancement of Computing in Education (AACE), P.O. Box 3728, Norfolk, VA 23514. Tel: 757-623-7588; e-mail: info@aace.org; Web site: http://www.aace.org/DL/.

PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)

EDRS PRICE EDRS Price MF01/PC01 Plus Postage.

DESCRIPTORS Computer Mediated Communication; Cooperative Programs; Educational Resources; *Electronic Libraries; *Information Sources; Internet; *Library Collections; Science Libraries; Science Materials; *Scientific Research; *Scientific and Technical Information; Scientists; Shared Resources and Services

IDENTIFIERS *Digital Collections; *Digital Information Services

ABSTRACT

The need for accurate and current scientific information in the fast paced Internet-aware world has prompted the scientific community to develop tools that reduce the scientist's time and effort to make digital information available to all interested parties. The availability of such tools has made the Internet a vast digital repository of information. But the ad hoc nature in which information is gathered and organized on the Web, makes access to such information a time consuming and sometimes frustrating affair. Digital library systems have the potential for solving problems in maintaining high quality scientific content delivered via the Web by providing tools for scientists to collect, verify, organize, manage, and update their collections. This paper describes an environment that reduces the effort and time required by scientists to share their data with other collaborators in an automated and asynchronous manner, thereby allowing them to focus mostly on their own scientific practice. The data is maintained as a collaborative collection in a digital library that can also be used as an educational resource. (Contains 22 references and 3 figures.) (Author)

Reproductions supplied by EDRS are the best that can be made from the original document.

ED 477 060

Enhancing Scientific Practice and Education Through Collaborative Digital Libraries

Gaurav Maini, John J. Leggett, Teongjoo Ong
Center for the Study of Digital Libraries, Texas A&M University, College Station, Texas, USA
Email: {gaurav, leggett, mong}@csdl.tamu.edu

Hugh D. Wilson, Monique D. Reed
Department of Biology, Texas A&M University, College Station, Texas, USA
Email: {wilson, monique}@mail.bio.tamu.edu

Stephan L. Hatch, John E. Dawson
Department of Rangeland Ecology and Management, Texas A&M University, College Station, Texas, USA
Email: s-hatch@tamu.edu; eddy@xmission.com

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

G.H. Marks

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Abstract: The need for accurate and current scientific information in the fast paced Internet-aware world has prompted the scientific community to develop tools that reduce the scientist's time and effort to make digital information available to all interested parties. The availability of such tools has made the Internet a vast digital repository of information. But the ad hoc nature in which information is gathered and organized on the web, makes access to such information a time consuming and sometimes frustrating affair. Digital library systems have the potential for solving problems in maintaining high quality scientific content delivered via the web by providing tools for scientists to collect, verify, organize, manage, and update their collections. This paper describes an environment that reduces the effort and time required by scientists to share their data with other collaborators in an automated and asynchronous manner, thereby allowing them to focus mostly on their own scientific practice. The data is maintained as a collaborative collection in a digital library that can also be used as an educational resource.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

1. Introduction

Digital library systems have the potential for solving problems in maintaining high quality scientific content delivered via the web by providing tools for scientists to collect, verify, organize, manage, and update their collections. The ability of the Internet to bring geographically distant scientists and researchers together as a community provides the opportunity to build collaborative digital libraries that can provide a larger quantity and higher quality of data. The notions of ownership and attribution of intellectual and scientific property must be retained and the contributors must control sharing of their data.

This paper describes an environment that reduces the effort and time required by scientists to share their data with other collaborators in an automated and asynchronous manner, thereby allowing them to focus mostly on their own scientific practice. The data is maintained as a collaborative collection in a digital library that can also be used as an educational resource [1]. The collection can be continually updated by the contributors and used or reused to build course content via tools provided by the digital library [2].

2. Research Context

2.1 Digital Flora of Texas

The Digital Flora of Texas (DFT) [3] project was established as an open, web-based digital resource. Participants currently include scientists from 43 contributing herbaria as well as many researchers at public

IR021788

BEST COPY AVAILABLE

and private organizations [4]. Each participating herbarium (or collection) has its own curator or group of curators who verify and validate the data they contribute. The DFT project has a wide array of web tools, collections and prototype systems that can be used for research or education. The DFT contains information visualization tools for displaying statistical and distributional maps of taxa of the flora of Texas. In addition, upload and data conversion tools provide an automated way for researchers to easily build collections and share them with the scientific community. The DFT collections are accessible through the web, making it easier for researchers, scientists, instructors and students to access accurate and up to date information. Instructors can use the collections to construct web-based content for their courses. These educational and scientific resources can be used and reused in many different contexts [5,6,7].

2.2 Collections

DFT collections are built either by a group of collaborating researchers or by individual researchers. In either case, the collections are publicly available and all scientists/researchers can participate in verifying and validating the result. Researchers, who are geographically distributed, collaborate using the various tools provided by the DFT project to build and maintain the following collections:

- The Herbarium Specimen Browser is the Internet's only multi-herbarium specimen data portal. It offers unique filtering, mapping and listing displays for a growing mass (more than 240,000 currently) of specimens taken from the participating Texas herbaria [8].
- The Image Gallery provides access to 8000 images of plants representing families found in Texas [9].
- The Contributor collection contains contact information for all contributors who generate digital content for this community-based collaborative digital library [10,11].

The following collections are a result of individual research:

- The Bibliography (Wilson) provides over 3000 bibliographic references for the Texas Flora [12].
- The Checklist of the Vascular Plants of Texas (Hatch, et al) refers to 180 families, ~1300 genera, and ~5000 species [13].
- The Centex Flora (Reed) is a manual of the dicot flora of Brazos and surrounding counties that provides keys and descriptions for 104 families, 489 genera and 1104 species [14,15].
- The Texas Grasses (Hatch and Dawson) has keys, descriptions and images for 142 genera and ~550 species in the grass family [16].

3. Design of a Collaborative Digital Library

3.1 Model of Collaboration/Interaction with System Components

In the design of a scientific collaborative digital library, it is of utmost importance for the contributor to maintain ownership of his/her scientific and intellectual property. The notion of ownership and control is maintained in the DFT because the contributors decide when to share their data and what part of their data to share. Automation is used only under control of the contributor and is a two step process: 1) uploading data to share and 2) incorporating data into the public collection. The sharing occurs at different times and from different geographic locations.

3.2 Rationale for Choice of Tools

The DFT is using the previously defined Flora of Texas Consortium (FTC) data model [17] as the common data format for its collections. eXtensible Markup Language (XML) [18] is ideal for the FTC data model since it renders a strong sense of structure to data and can be validated and maintained with little or no programming. A wide variety of Application Programming Interfaces (APIs) are also available that can convert a wide range of data formats to XML.

Greenstone is an open source digital library system that automatically provides multiple, user-defined browsing and searching indexes on the collections it maintains. It provides plugins for converting many standard format files (e.g. Word, PDF) to XML and allows rebuilding of collections while they are in use

(no down time). It also provides a standard, yet customizable user interface and compresses its documents, thereby saving space and, more importantly, mitigating the effect of long XML tags.

3.3 System Architecture

Contributors can upload their data and build or rebuild a collection in the DFT at any time. The contributors may use any program they wish to maintain their personal collections (e.g. Word, Excel, Access, Filemaker Pro). When they are ready to upload their data, they dump it in a flat ASCII file (an operation that is standard with all data management tools). The upload system converts the data into XML and validates it against the FTC data model. Errors, if any, are reported to the contributors. Records without errors are uploaded and are ready to be added to the collection. The uploaded data is not shared with the other collaborators until the contributor explicitly builds or rebuilds a collection of this validated data, making it available for public use. Other collaborators as well as non-specialists with an interest in the domain can access the collection using the common web-based user interface provided by Greenstone or by customizing their own user interface. This process relieves the scientists of the tasks of converting data into various acceptable formats enabling them to channel more time and effort toward practicing their science (Figure 1).

The Herbarium Specimen Browser, Image Gallery and Contributor collections mentioned in section 2 are collections that are built and maintained by geographically distributed contributors and follow the same methodology as described above. To upload his/her data, the contributor uses the web-based upload system [19] to specify the number and order of fields, field separators and field enclosures. Figures 2 and 3 show the upload form and an example search results page for the Image Gallery collection.

4. Lessons Learned

4.1 Scientific Practice

"The products of Systematic Botany, previously generated locally as static, hardcopy documents, can now be presented as collaborative enterprises from distributed centers as high-content, dynamic data resources that are constantly updated and refined." [1] The system architecture of the DFT enhances the sharing and browsing aspects of collaboration by providing a web-based input system that converts the uploaded data to standard XML format. This data is then added to the shared collection under control of the contributor. Thus, scientists can follow this process without disrupting their normal working style. This automated process shortens the time to share scientific information and the notion of ownership of intellectual and scientific property is maintained, as the scientists control their own collections and share whenever they are ready. The upload system also filters the data and performs various error checks (e.g. typographical error checks) in order to verify the data. Scientists appreciate the fact that their data is verified by the input system and by the information visualization tools (e.g. the mapping system) provided to browse through the digital library.

4.2 Education

We have observed that collaborators use and reuse either their own information content or that of the other collaborators in the digital library to build new course structures and educational packages. Tools provided with the digital library expedite the process of reuse and construction of new packages. Students are also presented with a larger and more diverse collection and are not restricted by the limitations of a localized collection. In most cases, we observed that collaborators found it useful to have cross collection querying/browsing, wherein the results of a search on one collection are used to search an entirely different collection. This exploratory learning methodology helps students find information they are looking for while following trails that interest them the most. Greenstone supports this cross collection querying/browsing model, enabling the user to seamlessly move from one collection to another.

4.3 Digital Libraries

The immediate benefit of having a collaborative digital library is the continuous updating and refinement of the information content and accuracy of information provided by the collaborative interactions of the scientists. Since it is fairly easy to maintain and browse the digital library, we find more and more collections being built using the system we have described. For example, the Image Gallery collection was developed because the scientists involved find the model of collaboration easy to use. Collections do indeed have their own unique identities and are a scientifically valid resource, but they also tend to become integrated in the notion of a digital library that allows cross collection querying/browsing.

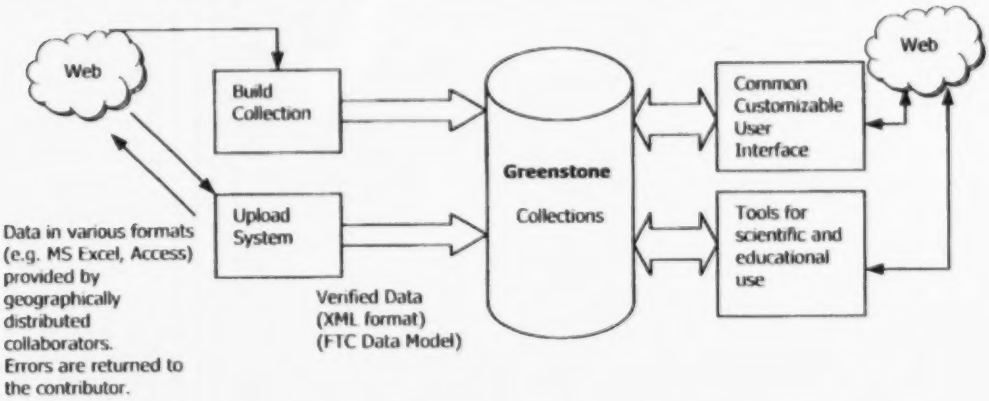
Web
Build Collection
Upload System
Data in various formats (e.g. MS Excel, Access) provided by geographically distributed collaborators. Errors are returned to the contributor.
Verified Data (XML format) (FTC Data Model)
Greenstone Collections
Common Customizable User Interface
Tools for scientific and educational use
Web

Figure 1. Architecture for DFT Collaborative Digital Library Collections.

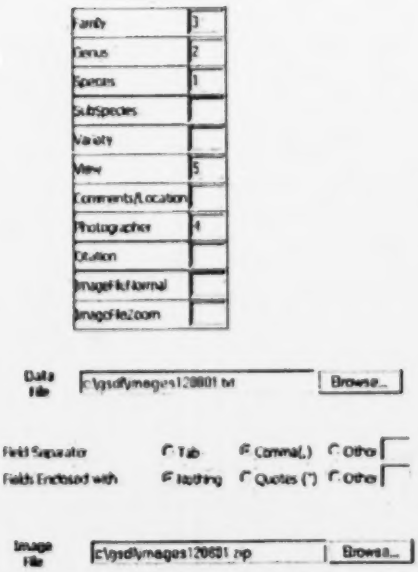

Figure 2. Example upload form for the Image Gallery Collection.

Figure 3. Example search results page for the Image Gallery Collection.

BEST COPY AVAILABLE

5. Future Work

Currently, the system does not have collaborative tools to support a peer review process by a group of curators [20]. With this set of tools, any person with botanical interests and a computer connected to the Internet could contribute to the digital library. The curators would be responsible for accepting or rejecting contributions and for maintaining high quality content. Considering the ease of creating and maintaining collections, we would like to put more content in the digital library in the form of a monthly journal, a collection of thesis and dissertations, and a collection of videos of lectures and field trips; thus adding various media formats to the digital collection and enhancing the value of the collections. Currently, collaborating scientists and university students are the main beneficiaries of the DFT. It would also be interesting to see how well the system scales to the needs of primary and secondary students.

References

[1] Hugh D. Wilson (2001). *Informatics: new media and paths of data flow*. Taxon 50: 381-387.

[2] Dipto Chakravarty, Jon Mason, Kevin Riley (2001). *IMS Digital Repositories White Paper* Revision: 22 August 2001

[3] http://www.texasflora.org/

[4] http://www.texasflora.org/dft_tx_herbs1.htm

[5] http://www.csdl.tamu.edu/FLORA/301Manhart/index.htm

[6] http://www.csdl.tamu.edu/FLORA/Wilson/481/s99/semhome99.htm

[7] http://www.csdl.tamu.edu/FLORA/fsb/fsbhome1.htm

[8] http://www.csdl.tamu.edu/FLORA/tracy2/main1.html

[9] http://www.csdl.tamu.edu/FLORA/ftc/dft/library.htm

[10] http://botany.cs.tamu.edu/contributor/

[11] http://botany.cs.tamu.edu/gsdl/cgi-bin/library?site=localhost&a=p&p=about&c=contributor&ct=0

[12] http://www.csdl.tamu.edu/FLORA/ftc/dft/bibbase.htm

[13] http://www.csdl.tamu.edu/FLORA/taes/tracy/home2.html

[14] http://www.csdl.tamu.edu/FLORA/centexflora/testing/demo1.html

[15] http://botany.cs.tamu.edu/gsdl/cgi-bin/library?site=localhost&a=p&p=about&c=centexflora&ct=0

[16] http://www.csdl.tamu.edu/FLORA/taes/tracy/610/index.html

[17] http://www.csdl.tamu.edu/FLORA/ftc/ftcffld4.htm

[18] http://www.w3.org/XML/

[19] http://botany.cs.tamu.edu/dfot

[20] Goh, D., and Leggett, J.J. (2000). *Patron augmented digital libraries.* In *Proceedings of the Digital Libraries '00 Conference*, San Antonio, TX, pp. 153-163, June 2-7, 2000.

[21] http://www.csdl.tamu.edu/FLORA/tamuherb.htm

[22] Erich R. Schneider, John J. Leggett, Richard K. Furuta, Hugh D. Wilson, Stephan L. Hatch (1997). *WWW tools for accessing botanical collections.* In *Proceedings of WebNet 97*, November 1-5, 1997, Toronto, Ontario. (Winner: Best Student Paper Award.)

Acknowledgements

This research was supported in part by the Texas Advanced Research Program under Grant Number 010366-0041c-1999.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").